EEG-based emotion classification using graph signal processing

Seyed Saman Saboksayr

Dept. of Electrical and Computer Engineering

University of Rochester

Email: ssaboksa@ur.rochester.edu

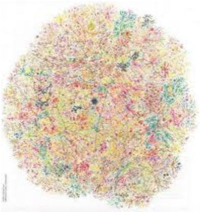
Co-authors: Gonzalo Mateos and Mujdat Cetin

Acknowledgement: NSF awards CCF-1750428, CCF-1934962 and ECCS-1809356

IEEE ICASSP, June 6-11, 2021

Online social media

Internet



Clean energy and grid analytics

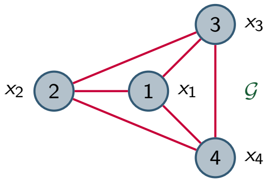
- Network as graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$: encode pairwise relationships
- Desiderata: Process, analyze and learn from network data [Kolaczyk'09]

  $\Rightarrow$ Use $\mathcal{G}$ to study graph signals, data associated with nodes in $\mathcal{V}$
- Ex: Opinion profile, buffer congestion levels, functional brain connectivity
- **Q:** What about classification data using learned graphs?

- Undirected $\mathcal{G}$ with adjacency matrix $\mathbf{W}$ $\Rightarrow W_{ij}$ = Proximity between i and j
- Define a signal $\mathbf{x}$ on top of the graph $\Rightarrow x_i$ = Signal value at node i
- Associated with $\mathcal{G}$ is the combinatorial graph Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{W}$

  $\Rightarrow$ $\mathbf{D}$ is diagonal matrix where its elements are vertices degree

  $\Rightarrow$ Graph Laplacian is positive semidefinite so $\mathbf{L} = \mathbf{V\Lambda V}^\top$



- Graph Signal Processing $\rightarrow$ Exploit structure encoded in $\mathbf{W}$ or $\mathbf{L}$ to process $\mathbf{x}$

  $\Rightarrow$ Use GSP to learn the underlying $\mathcal{G}$ or a meaningful network model

Graph Fourier transform (GFT) and signal smoothness

- Graph Fourier transform (GFT): $\tilde{\mathbf{x}} := \mathbf{V}^\top \mathbf{x}$ for undirected graphs
  - $\Rightarrow$ Decompose $\mathbf{x}$ into different modes of variation
  - $\Rightarrow$ Inverse (i)GFT $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$, eigenvectors as frequency atoms
- Total variation (or *Dirichlet energy*) of signal $\mathbf{x}$ with respect to $\mathbf{L}$

$$\mathrm{TV}(\mathbf{x}) := \mathbf{x}^\top \mathbf{L}\mathbf{x} = \frac{1}{2}\sum_{i \neq j} W_{ij}\left(x_i - x_j\right)^2 = \sum_{k=1}^{N} \lambda_k \tilde{x}_k^2$$

  - $\Rightarrow$ Smoothness measure on the graph $\mathcal{G}$
- For Laplacian eigenvectors $\mathbf{V} := [\mathbf{v}_1, \ldots, \mathbf{v}_N] \Rightarrow \mathrm{TV}(\mathbf{v}_k) = \lambda_k$
  - $\Rightarrow 0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_N$ can be viewed as frequencies

- labeled graph signals $\mathcal{X}_c := \{\mathbf{x}_p^{(c)}\}_{p=1}^{P_c}$ from C different classes.

  $\Rightarrow$ Signals in each class possess a very distinctive structure

- **Assumption:** Class c signals are smooth w.r.t. unknown $\mathcal{G}_c(\mathcal{V}, \mathcal{E}_c)$

- Multiple linear subspace model

  $\Rightarrow$ Signals spanned by few Laplacian modes (GFT components)

  $\Rightarrow$ Like susbpace clustering [Vidal'11], but with supervision

Problem statement

Given training signals $\mathcal{X} = \bigcup_{c=1}^{C} \mathcal{X}_c$, learn discriminative graphs $\mathbf{W}_c$ under smoothness priors to classify test signals via GFT projections.

Topology inference: Motivation and context

- Network topology inference from nodal observations [Kolaczyk'09]
  - Partial correlations and conditional dependence [Dempster'74]
  - Sparsity [Friedman et al'07] and consistency [Meinshausen-Buhlmann'06]
- Key in neuroscience [Kassam et al'13] $\Rightarrow$ Functional network from EEG signal
- Noteworthy GSP-based approaches
  - Graphical models [Egilmez et al'16], [Rabbat'17], [Kumar et al'19], ...
  - Smooth signals [Dong et al'15], [Kalofolias'16], [Sardellitti et al'17], ...
  - Streaming data [Shafipour et al'18], [Natali et al'20], [Saboksayr et al'21], ...
- **Our contribution:** learn a discriminative graph-based representation of smooth signals

- Given a set of graph signal observations from P trials $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_P] \in \mathbb{R}^{N \times P}$

  $\Rightarrow$ A link between smoothness and sparsity (considering $\mathbf{Z} \in \mathbb{R}_+^{N \times N}$ as Euclidean-distance matrix)

$$\sum_{p=1}^{P} \mathrm{TV}(\mathbf{x}_p) = \mathrm{trace}(\mathbf{X}^\top \mathbf{L} \mathbf{X}) = \frac{1}{2} \|\mathbf{W} \circ \mathbf{Z}\|_1$$

- Framework for learning graphs under a smoothness prior [Kalofolias'16]

$$\min_{\mathbf{W}} \|\mathbf{W} \circ \mathbf{Z}\|_1 + f(\mathbf{W}) \quad \text{s. t.} \quad \mathrm{diag}(\mathbf{W}) = \mathbf{0},\ W_{ij} = W_{ji} \geq 0,\ i \neq j$$

  $\Rightarrow$ Convex objective function $f(\mathbf{W})$ encodes assumptions about the network $\mathcal{G}$

  $\Rightarrow$ Primal-dual solver amenable to parallelization, $\mathcal{O}(N^2)$ cost [Kalofolias'16], [Komodakis et al'15]

- Discriminative graph learning per class c

$$\min_{\mathbf{W}_c \in \mathcal{W}_m} \|\mathbf{W}_c \circ \mathbf{Z}_c\|_1 - \alpha \mathbf{1}^\top \log(\mathbf{W}_c \mathbf{1}) + \beta \|\mathbf{W}_c\|_F^2 - \gamma \sum_{k \neq c}^{C} \|\mathbf{W}_c \circ \mathbf{Z}_k\|_1$$

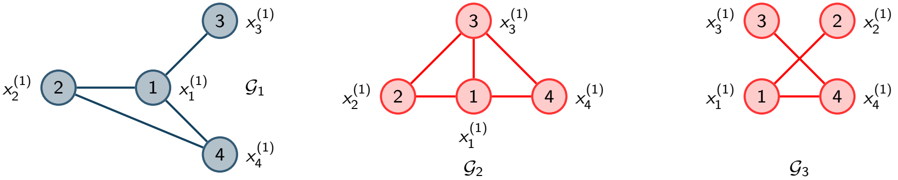
$$\mathcal{W}_m = \left\{ \mathbf{W} \in \mathbb{R}_+^{N \times N} : \mathbf{W} = \mathbf{W}^\top, \operatorname{diag}(\mathbf{W}) = 0 \right\}$$

$\Rightarrow$ Capture the underlying graph topology (class c structure)

$\Rightarrow$ Discriminability to boost classification performance

$\Rightarrow$ Logarithmic barrier forces positive degrees

$\Rightarrow$ Penalize large edge-weights to control sparsity



- **Q:** Given graphs $\{\hat{\mathbf{W}}_c\}_{c=1}^{C}$, how do we classify a test signal $\mathbf{x}$?

  $\Rightarrow$ Pass $\mathbf{x}$ through a filter-bank with C low-pass filters

$$\tilde{\mathbf{x}}_{F,c} = \text{diag}(\tilde{\mathbf{h}})\hat{\mathbf{V}}_c^{\top}\mathbf{x} \quad \Longrightarrow \quad \hat{c} = \underset{c}{\text{argmax}}\left\{\|\tilde{\mathbf{x}}_{F,c}\|_2^2\right\}$$

  $\Rightarrow$ $\tilde{\mathbf{h}} = [\tilde{h}_1, \ldots, \tilde{h}_N]^{\top}$ is the frequency response of an ideal low-pass filter

  $\Rightarrow$ Learned class-c GFT basis $\hat{\mathbf{V}}_c$

If $\mathbf{x}$ belongs to class $c^{\star}$, say, then this graph signal should be smoothest with respect to $\mathcal{G}_{c^{\star}}$. Equivalently, for fixed low bandwidth we expect the signal power to be largest when projected onto the GFT basis constructed from $\hat{\mathbf{L}}_c$.

- Emotion is an important aspect of human life
  - ⇒ Brain pattern associated with a specific emotion
- Why EEG? non-invasive, affordable, high temporal resolution
- **DEAP data-set:** [Koelstra et al'11]
  - ⇒ 32 participants rated 40 one-minute-long music video
  - ⇒ Signals acquired from $N = 32$ EEG channels
  - ⇒ Rating levels of valence and arousal from 1 to 9

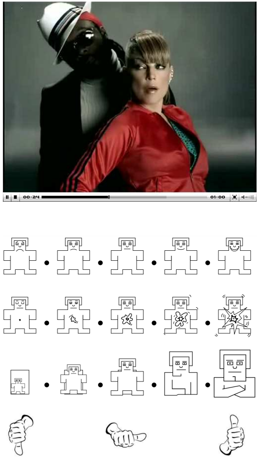


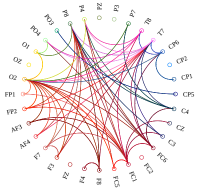
Figure extracted from [Koelstra et al'11]

- We perform a subject-specific valence and arousal classification task
  - $\Rightarrow$ low valence/arousal (rating < 5) and high valence/arousal (rating ≥ 5)
  - $\Rightarrow$ Learn $C = 2$ graphs and project onto the 8 smoothest modes
  - $\Rightarrow$ Report leave-one (trial)-out classification accuracy

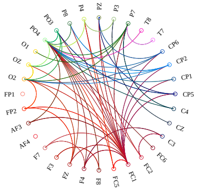
Study	Valence	Arousal
Proposed method	**92.73**	**93.44**
[Kalofolias'16]	86.56	88.91
[Chao et al'20]	77.02	76.13
[Rozgić et al'13]	76.90	69.10
[Chen et al'15]	76.17	73.59
[Tripathi et al'17]	81.40	73.36

- Established superior performance $\Rightarrow$ Any discriminative pattern in the underlying learned graph?

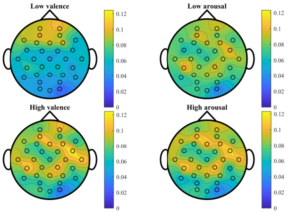
- Learn two graphs corresponding to low and high emotions per person
  - ⇒ We do classification via low-frequency components
    - → Mapping by low-frequency eigenvectors
- The asymmetrical pattern of the frontal EEG activity [Schmidt et al'01]
- Different patterns in: 1. left frontal 2. right temporal 3. parietal [Nie et al'11]



Different connections ($p \leq 0.002$) for valence



Different connections ($p \leq 0.03$) for arousal



8 smoothest modes

Closing remarks

- Discriminative graph learning
  - ⇒ Signals are smooth w.r.t. unknown class-specific graphs $\mathcal{G}_c$
  - ⇒ Discriminability via enforcing non-smoothness
  - ⇒ Classification by low-pass filtering
- EEG emotion recognition
  - ⇒ Outperform state-of-the-art methods in classifying low and high valence/arousal
  - ⇒ Infer discriminative brain patterns
- Ongoing work
  - ⇒ Proximal gradient based and Nesterov-type accelerated algorithms
  - ⇒ Observations of streaming signals

Extended journal version is available at: https://arxiv.org/abs/2101.00184